

# Exploitation de l'hétérogénéité entre points de vue-opinion

## Une approche par l'alignement

**Samuel GESCHE**

LIRIS – INSA de Lyon  
20, avenue Albert Einstein  
69621 VILLEURBANNE CEDEX  
samuel.gesche@insa-lyon.fr

---

*RÉSUMÉ. La confrontation de points de vue scientifiques est une activité vieille comme la science. Cependant, jamais les scientifiques n'avaient été confrontés à un tel volume de publication qu'aujourd'hui. Cette explosion documentaire peut mettre en péril toute une partie de la créativité des chercheurs, en étouffant leur capacité à tisser des liens entre leurs travaux et ceux de leurs confrères. Nous examinons dans cet article l'apport que peuvent fournir les techniques d'alignement d'ontologies, dans le cadre de l'élaboration d'un outil de support à la confrontation de points de vue scientifiques. Nous présentons également l'environnement que nous développons dans ce sens, ainsi que deux mises en application dans des domaines différents.*

*ABSTRACT. Confronting scholarly viewpoints is as old as science itself. And never the amount of published scholarly knowledge has been quite as huge as nowadays. This fast growth can eventually overwhelm the capacity of the scholars using them as a support for creativity, by diluting their ability to link their subject of research with other works. In our work, we study the concept of ontology matching, trying to build a tool for assisting scholarly viewpoints confrontation. We also present our prototype of such a tool, and its application to two different cases in two different domains.*

*MOTS-CLÉS: confrontation, alignement, point de vue, opinion, assisté par ordinateur*

*KEYWORDS: confrontation, alignment, viewpoint, opinion, computer-assisted*

---

## 1. Introduction

On considère généralement que l'hétérogénéité dans les conceptions humaines est une richesse appréciable. La science est un domaine dans lequel cette hétérogénéité est essentielle par définition : la créativité dans la recherche vient en grande partie de la confrontation des idées. De plus, cette hétérogénéité est parfaitement inéluctable : on se situe à un très haut niveau d'expertise. Il est donc impossible d'invoquer un arbitre pour trancher entre deux conceptions qui résistent à la preuve logique et à l'expérience. L'évaluation par les pairs est le seul outil utilisé actuellement. Or, il ne permet pas de juger de la validité d'une assertion. Il permet seulement de décider si elle mérite d'être diffusée (Piolat *et al.*, 2004).

On considère généralement que la numérisation des connaissances scientifiques est une pratique souhaitable. Les documents numérisés peuvent être mis en ligne dans des bibliothèques numériques (comme le site [cefael.efa.gr](http://cefael.efa.gr)). Ces bibliothèques peuvent être dotées de systèmes de recherche beaucoup plus performants que leurs homologues physiques. Certains systèmes permettent même aux experts d'interagir avec leur contenu, par le biais d'annotations par exemple (Bénel *et al.*, 2002). C'est d'autant plus intéressant que la quantité des connaissances scientifiques augmente à un point tel qu'il est illusoire d'espérer les réunir toutes dans une même bibliothèque universitaire physique. Dans certains domaines de recherche, il devient impossible de se tenir au courant de l'état de son art tout en faisant ses propres travaux de recherche. C'est la cause directe de travaux tels que (Aussenac-Gilles, 2006).

On considère généralement que l'hétérogénéité dans les données numériques est une mauvaise chose. L'hétérogénéité technique est une plaie ouverte dans le développement du tout-interopérable, l'hétérogénéité fonctionnelle en est une autre dans l'utilisation des logiciels, et l'hétérogénéité terminologique en est une troisième dans la communication entre systèmes. Nous pourrions aussi citer les incohérences dans les données. Or tout cela est précisément le reflet de cette richesse appréciable et le résultat inéluctable de cette numérisation souhaitable. En somme, la science numérique, qui fait l'éloge de l'hétérogénéité, est opposée à la manufacture numérique, qui fait celui de l'homogénéité.

Ainsi donc, l'hétérogénéité scientifique est bonne, la numérisation de cette hétérogénéité est bonne, mais l'hétérogénéité numérique est mauvaise. Pourtant, le besoin d'exploiter l'hétérogénéité de la production scientifique par l'outil numérique se fait de plus en plus impérieux. En effet, l'accroissement exponentiel de la production scientifique peut à terme aboutir à une dissolution des capacités de lecture et donc de la confrontation des idées par le support écrit. Il est donc à espérer que tout cela soit renforcé par des moyens numériques dans les prochaines années. L'alignement d'ontologies, qui est utilisé en particulier dans le domaine de l'interconnexion des services Web, peut jouer un rôle moteur dans cette exploitation de l'hétérogénéité. Il est cependant nécessaire de réduire l'habituel fossé sémantique, particulièrement étendu dans le cadre de la production scientifique.

Nous proposons donc de vous présenter nos recherches concernant la confrontation assistée par ordinateur de points de vue-opinion. Nous commencerons par présenter notre approche de manière globale, en expliquant comment nous utilisons les techniques développées pour l'alignement d'ontologies dans le contexte de deux cas d'étude, puis nous détaillerons la manière dont nous intégrons ces techniques dans un outil amené à répondre à la problématique que nous avons présentée : soutenir la créativité scientifique par une assistance numérique.

## **2. Contexte**

### **2.1. Points de vue-opinion**

Il est d'abord important de préciser la manière dont nous concevons les points de vue-opinion.

Tout d'abord, les points de vue que nous traitons sont des points de vue d'experts. Nous incluons dans le terme d'expert le chercheur et le scientifique, qui sont des experts particuliers.

Nous appelons « point de vue-opinion », pour « point de vue de type opinion », la conception qu'un expert a d'une problématique donnée. Cette conception est soit personnelle, soit partagée dans le cadre d'une équipe de recherche restreinte. C'est un objet au moins en partie mental, cumulant aussi bien un positionnement au niveau de la discipline (point de vue) qu'un positionnement personnel (de type opinion). Pour autant, il ne s'agit pas d'une conception floue et incohérente. Simplement, elle est incomplète sans certains non-dits et certains inter-dits (Iacovella *et al.*, 2006).

Nous appelons « point de vue exprimé » un document faisant état d'un point de vue. Que ce soit une publication, un rapport de recherche interne ou simplement un document de travail ou une partie d'un cahier de laboratoire, il traite du point de vue sous une forme qui peut être partagée avec des personnes extérieures. C'est un objet strictement documentaire.

Nous appelons « point de vue modélisé » un point de vue exprimé dans un formalisme de modélisation. Ce document, en plus de permettre le partage, réduit l'incertitude quant à son interprétation.

Enfin, nous appelons « point de vue sous forme de graphe » un point de vue exprimé dans un formalisme de représentation par graphe. Comme ce type de modèle est particulièrement facile à traiter dans le cadre d'une analyse numérique, un bon nombre de solutions de traitement numérique des points de vue ou des opinions utilisent des représentations par graphes (Aussenac-Gilles, 2006, El Asri *et al.*, 2005, Lahna *et al.*, 2005).

Le choix du degré de formalisation d'un point de vue-opinion est celui d'un compromis entre l'exhaustivité d'un côté, la cohérence et la simplicité de l'autre. Le

point de vue mental est le plus complet, tandis que le point de vue sous forme de graphe est le plus facile à comprendre, du moins pour une machine.

## **2.2. Confrontation de points de vue-opinion**

La confrontation de points de vue-opinion est tout simplement la manière dont les experts travaillent avec différents avis ou différentes tendances dans la même discipline. Il s'agit d'une activité à part entière de la recherche. Elle est source de créativité et d'innovation grâce aux liens que l'on peut construire entre les points de vue ; elle est également partage de connaissance par les ouvrages et synthèse qu'elle permet d'écrire. Et elle est remise en question pour celui qui fait face à des avis contraires.

Nous appelons « confrontation assistée par ordinateur » un processus de confrontation faisant appel à la machine dans un certain nombre de ses composantes, mais toujours soumis à la direction et à la validation de l'utilisateur humain. C'est l'approche que nous avons choisie. Elle consiste à fournir un plan de travail numérique permettant la visualisation des points de vue, puis leur confrontation. Ces activités sont outillées par des fonctionnalités pouvant être utilisées indépendamment les unes des autres.

## **2.3. Cas d'Etude**

Cette problématique de confrontation n'est pas arrivée par hasard. Il s'agit d'une facette d'une activité de recherche concernant l'impact du paradigme numérique sur le travail de l'expert (Bénel, 2006). La plateforme Porphyry ([www.porphyry.org](http://www.porphyry.org)), développée dans le cadre de cette activité de recherche, permet la confrontation de points de vue portant sur un même corpus documentaire. Nous avons envisagé la confrontation dans un cadre plus général.

Ainsi, nous appuierons notre propos sur deux projets différents. Le premier est Porphyry, que nous venons d'évoquer ; le second est Towntology ([www.towntology.net](http://www.towntology.net)), un projet concernant la conception d'ontologies urbaines. Ces deux projets ont donné naissance aux plateformes portant les mêmes noms, qui permettent toutes deux à des experts de construire des points de vue sous forme de graphes, par le biais d'une interface appropriée.

### **2.3.1. Porphyry**

La plateforme Porphyry est issue de l'atelier de recherche transdisciplinaire ARTCADHi, lequel s'intéresse au processus de construction du sens au sein des sciences humaines et sociales. C'est une plateforme bâtie sur une architecture multi-tiers et un client lourd. A l'heure actuelle, plusieurs serveurs sont en exploitation, à destination de différentes équipes de recherche en France.

Etant donnée les contraintes posées par l'architecture distribuée de Porphyry, nous nous appuyons sur un corpus de points de vue extérieur aux serveurs. Ce corpus a été élaboré il y a trois ans à partir du Bulletin de Correspondance Hellénique de l'Ecole Française d'Athènes, mis en ligne sur le site [cefael.efa.gr](http://cefael.efa.gr). Il comporte le sommaire de la rubrique Chronique des Fouilles, et présente donc l'évolution de la cartographie archéologique du monde hellénique au cours du vingtième siècle. En tout, nous disposons là de dix points de vue.

### 2.3.2. *Towntology*

Le projet *Towntology* est né de la volonté de construire des ontologies des domaines liés à l'urbanisme. En effet, l'urbanisme n'a pas une terminologie figée, ni dans l'espace, ni entre acteurs. La plateforme développée dans le cadre du projet permet donc la construction d'ontologies pré-consensuelles (Keita *et al.*, 2006), en vue de formaliser les points de vue des experts sur leur discipline.

Le projet *Towntology* apporte ses propres contraintes : les ontologies sont bâties dans le cadre multilingue d'une action COST ; l'urbanisme est par ailleurs une discipline très générale recouvrant de multiples entités cohérentes telles que les transports, la voirie, les services municipaux etc. Pour notre étude, nous disposons de trois ontologies en français décrivant trois domaines différents (la voirie, les transports et la ville).

## 3. Alignement et Confrontation

### 3.1. *Ontologies légères et lourdes*

Les ontologies ont été développées il y a bientôt vingt ans. Leur objectif est, comme l'a formulé (Gruber, 1993), de spécifier de manière formelle une conceptualisation sémantique. Elles sont, depuis ce temps, devenues un standard de fait dans la représentation des connaissances, et elles sont un pilier du Web Sémantique.

Une ontologie légère est composée d'une hiérarchie de concepts, d'une hiérarchie de relations, et d'un ensemble de connexions entre concepts et relations. Il s'agit de la structure la plus simple qui soit appelée « ontologie ». Elle fournit cependant un niveau de formalisation acceptable pour un certain nombre d'applications. Par exemple, les ontologies pré-consensuelles de *Towntology* sont des ontologies légères.

Dans certains cas, les ontologies légères ne fournissent pas un support suffisamment formel pour le raisonnement. Il s'agit en particulier des ontologies de domaine et de résolution utilisées dans les systèmes experts (Breuker *et al.*, 1994). On utilise alors, depuis le début du siècle, des ontologies lourdes. Ces ontologies sont enrichies d'un certain nombre de primitives, telles que des cardinalités liées aux relations et des axiomes permettant d'utiliser des mécanismes d'inférence.

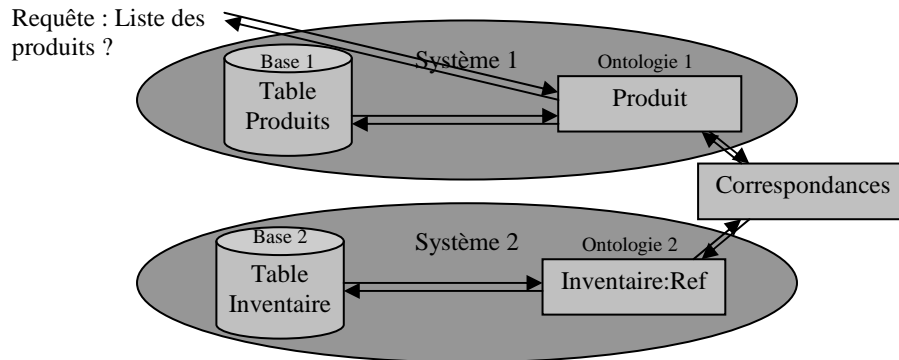
### 3.2. Alignement d'ontologies

Quand les ontologies se sont développées, avant de s'imposer comme un standard dans la représentation des connaissances, la tentation a été grande d'en faire la « panacée » de la gestion des connaissances. En somme, les ontologies devaient réussir là où les systèmes experts avaient échoué. Plus entreprenants encore, certaines équipes ont tenté de bâtir une ontologie générale de tout ce qui existe (Guha *et al.*, 1990), essayant ainsi de décrire l'*ontologie* au sens philosophique.

Las, si les ontologies ont effectivement révolutionné le monde des systèmes experts, elles ont en règle générale échoué à concrétiser le rêve de l'humanité pour lesquelles elles étaient nées, à savoir maîtriser la connaissance humaine dans son intégralité. Le constat a été sans appel au début du vingt-et-unième siècle, et la formulation en est devenue célèbre dans le milieu des ontologies : l'utilisation des ontologies ne réduit pas l'hétérogénéité, tout juste permet-elle de la refouler au niveau supérieur (Euzenat, 2007). Si deux personnes ont une conception différente d'un même problème, ce n'est pas la construction d'ontologies qui va les réconcilier : on obtiendra deux ontologies différentes et probablement incompatibles.

C'est dans ce contexte qu'est apparu, depuis quelques années, l'alignement d'ontologies. La philosophie de l'alignement est relativement simple : si les ontologies sont hétérogènes, elles peuvent cependant être interopérables. Il suffit de faire un « dictionnaire » permettant de passer de l'une à l'autre. Ainsi l'alignement vise-t-il à produire une table de correspondance entre des éléments des ontologies, comme on peut le voir par exemple sur la figure 1. On peut ainsi interfacer deux systèmes conçus avec des préoccupations différentes avec une même application de recherche (dans cet exemple, on recherchera une liste des produits aussi bien dans une première base contenant une table *produits* que dans une seconde ne contenant pas une telle table, en sachant que ces deux bases sont décrites dans des ontologies : on inscrira dans la table de correspondance que le *produit* de l'un est le champ *ref* de la table *inventaire* de l'autre).

Etant donné le degré de formalisation des ontologies, une quantité assez impressionnante d'outils est à la disposition de celui qui veut élaborer un algorithme d'alignement. Cela peut être aussi simple que la comparaison terminologique, avec ses multiples algorithmes (certains déjà fort anciens, l'un des plus connus étant probablement la distance de Levenshtein (Levenshtein, 1966)). Mais la plupart du temps, on tire parti de la structure de l'ontologie, en utilisant des techniques basées sur les relations locales, les attributs et propriétés des entités, les cardinalités, en allant jusqu'à considérer le métamodèle s'il est disponible, ou utiliser des logiques de premier ordre. Qui plus est, la combinaison de plusieurs de ces techniques se révèle souvent bien plus puissante que l'utilisation d'une méthode unique. Nous ne détaillerons pas ici toutes ces techniques, par ailleurs fort bien expliquées dans (Euzenat, 2007). Il est cependant intéressant de noter que l'on construit désormais des architectures d'alignement génériques, qui permettent de choisir et combiner à l'infini les différentes méthodes disponibles (Huzella *et al.*, 2007).



**Figure 1.** Exemple d'application de l'alignement d'ontologies à l'interconnexion de deux systèmes d'information commerciaux

Même si l'alignement se fait l'apôtre du consensus et de l'interopérabilité, son utilité est manifeste même si l'on se préoccupe de profiter de l'hétérogénéité. En effet, tout n'est pas inconciliable dans deux points de vue. Deux diplomates commencent par établir un langage commun avant de parler de ce qui les sépare. Il est donc intéressant de commencer par concilier ce qui peut l'être avant de passer aux vraies divergences.

### 3.3. Confrontation de points de vue-opinion

La confrontation de points de vue-opinion peut donc profiter des techniques d'alignement pour effectuer un premier dégrossissage. Cependant, l'alignement a un certain nombre de limites intrinsèques qui le rendent insuffisant pour effectuer la totalité de la confrontation :

- Tout d'abord, l'alignement tire parti du caractère formel des langages utilisés. On se rend rapidement compte, en étudiant les différentes techniques développées, combien il est profitable de disposer d'un formalisme aussi contraignant que possible. En tout état de cause, ce qui ne peut être exprimé par une primitive du formalisme est sujet à interprétation, et la machine ne sait pas interpréter. Dans notre cas, les experts, non informaticiens, ont besoin de formalismes laissant une grande place à l'interprétation. Les points de vue sont exprimés pour des humains dotés de sens commun, et non pour des calculateurs.

- Ensuite, l'alignement est souvent basé sur des lexiques, thésauri, voire même sur des ontologies du domaine qui sont indépendantes des structures alignées. La raison d'être en est simple : une terminologie non figée est un poison pour

l'alignement. Malheureusement, nos experts débattent sur leur expertise, et donc ils se débattent dans des problèmes terminologiques sans fin. Le mot qu'un archéologue utilise pour nommer le haut du bâtiment, l'autre l'utilise pour le bas. En urbanisme, le plan de la ville d'un service donné ne contient pas les mêmes rues que celui du service voisin. Aucun consensus terminologique ne peut être imposé, parce que la terminologie elle-même est un enjeu de confrontation.

– L'alignement est généralement conçu comme un processus binaire : on aligne deux structures, et non pas trois ou plus. Nous devons, quant à nous, envisager la confrontation de trois, cinq, voire dix points de vue.

– Enfin, et c'est le plus ennuyeux, l'alignement s'attache à définir des équivalences ou des partitions. Parfois, on peut voir des méthodes incluant la notion de différence. L'alignement, en tant que processus de calcul, est indéfectiblement lié à la logique. Le rapprochement de motifs identiques est certes nécessaire à la confrontation, mais en aucun cas suffisant. La confrontation a besoin de la possibilité de rapprocher des termes qui s'évoquent l'un l'autre (feu et pompiers, par exemple). Ou bien des termes liés par un lien sémantique (*AutoCAD est un logiciel pour la conception de automobile*). L'alignement est rarement envisagé comme la constitution d'une ontologie entre ontologies.

Toutes ces limitations ne diminuent en rien l'apport de l'alignement à la problématique de la confrontation de points de vue scientifiques, mais l'alignement ne peut pas être considéré comme « *la* » solution. Nous l'avons donc inclus au sein d'un environnement apte à combler ses insuffisances.

#### **4. Graphes et Documents**

Le document est l'incarnation du point de vue dans le monde des choses tangibles. Nous ne pouvons pas, faute d'avoir la capacité de télépathie, accéder au point de vue réel d'une personne. Seul peut en être capturé ce que cette personne va en exprimer, et seul peut être réellement utilisé ce qui en sera enregistré sur un support pérenne, plus fidèle en tous cas que la mémoire des auditeurs. Dans ce contexte, alors que nous nous intéressons réellement à la confrontation de points de vue –les conceptions des chercheurs– nous sommes bien obligés de nous contenter des documents que ces chercheurs ont rédigés.

Si cela était la seule contrainte, nous n'écririons pas cet article, parce que le tour de la question aurait été vite fait. Mais nous ne voulons pas nous-mêmes confronter des écrits. Nous voulons que ce soit un expert du domaine qui le fasse –seule garantie que la confrontation ait un résultat qui vaille la peine d'être noté. Et qui plus est, nous voulons équiper cet expert d'un outil d'assistance.

Or, force est de constater que les outils informatiques ne sont à l'aise qu'avec des structures de données qui finissent toujours pas inclure un (et souvent des) formalisme(s) de graphes. Tout simplement parce que le processeur calcule, et que le graphe est un excellent support de calcul : que ce soit pour la visualisation (un bon



dessin vaut mieux qu'un long discours, dit-on), pour le stockage (XML et les bases de données relationnelles en sont deux exemples) ou pour les processus de traitement (la Standard Template Library (STL) en C++ n'existe pas en l'absence de graphes). L'alignement ne fait pas exception : ce que l'on aligne, ce sont encore et toujours des graphes.

Mais voilà, l'être humain ne s'exprime pas naturellement par graphes. L'enregistrement d'une communication humaine est une séquence ordonnée, chaque élément suivant le précédent. Cela n'a pas changé avec l'arrivée de l'informatique : même si le stockage est parfois fait sous une forme graphique (en XML par exemple), il est rare que ce soit le discours, au sens pragmatique du terme, qui soit sous forme de graphe. Nous offrons donc la possibilité à l'expert utilisateur, *via* les fonctionnalités de désignation et d'organisation que nous détaillerons en section 5.2, d'apposer une structure de graphe sur un document linéaire. Cela permet ensuite d'utiliser la puissance de calcul de la machine de manière beaucoup plus efficace que par une simple indexation, parce que la structure formée est imprégnée de l'expertise de l'utilisateur et donc sémantiquement riche.

## **5. Modèle d'Interaction**

### **5.1. Approche Utilisateur**

En mettant l'expert au centre du procédé, nous recherchons à le faire profiter des capacités de calcul de l'informatique, mais sans pour autant prendre la main quand il ne le souhaite pas. Nous avons donc une approche de type « Assistée par Ordinateur ». Ainsi, nous fournissons une palette d'outils à l'expert qu'il emploie pour travailler sur un corpus de points de vue. Nous refusons pour l'instant tout processus séquentiel, d'une part parce que nous n'avons aucune raison d'avoir un protocole séquentiel de confrontation de points de vue, et ensuite parce qu'il est toujours possible de définir soi-même, en tant qu'utilisateur, une telle séquence. Il est cependant possible que l'étude des traces d'exécution nous montre quelques invariants.

Notre objectif général est de générer de nouvelles connaissances pour l'expert. Pour cela, il est intéressant qu'il soit intégré dans l'action autant que possible, afin que son intuition puisse guider sa recherche. A l'étude systématique des points de vue nous préférons par conséquent une exploration plus ludique, en somme un vagabondage propice à faire travailler l'imagination. Les outils sont à disposition de l'utilisateur, et il peut les utiliser quand bon lui semble, ne serait-ce que pour voir ce que cela fait. Une fonctionnalité de capture de zones de l'écran permet de constituer un portfolio qui, à terme, peut contenir plus de connaissances que la sauvegarde de la confrontation elle-même. Tout dans l'application peut être capturé, excepté (pour des raisons techniques principalement) les messages bloquants et les fenêtres contextuelles.

## 5.2. Interaction Homme-Machine

Notre approche de la confrontation place l'expert au cœur du processus. Ainsi, l'une de nos préoccupations est de fournir une interface aussi performante que possible pour une interaction homme-machine fertile.

Les modèles que nous étudions pour le moment sont tous des graphes, au sens informatique du terme (des nœuds reliés par des liens). Les formalismes varient, mais ils se rapprochent plus ou moins des réseaux sémantiques. Ainsi, la navigation dans ces graphes s'est rapidement imposée comme la méthode la plus efficace de parcourir les données. Notre application dispose donc d'une interface de navigation raisonnablement performante. Il est à noter que les graphes fabriqués par des experts non informaticiens n'ont généralement pas la facilité d'exploitation des arbres binaires. Ils se rapprochent plutôt, la plupart du temps, de séries d'énumérations, le cas extrême étant atteint lorsque les experts mettent tous les nœuds à la racine avant de les ordonner (Bénel, 2006).

Nous avons déterminé cinq grandes opérations lors de la confrontation de points de vue : l'observation, la désignation, l'organisation, la connexion et la séparation.

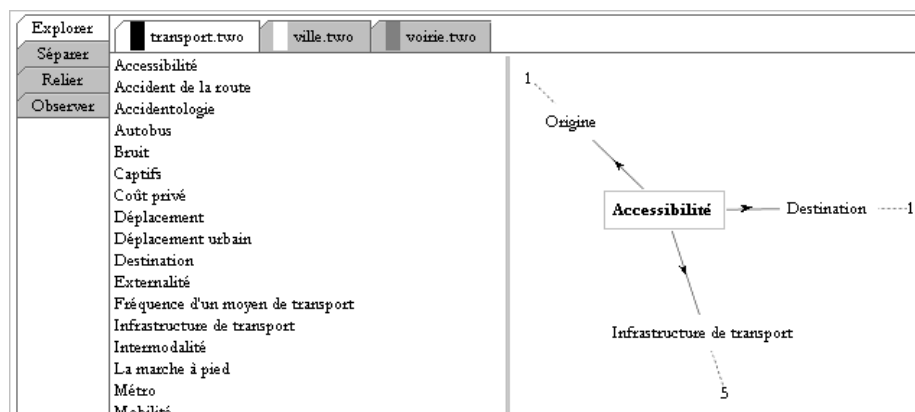


Figure 2. Observation initiale d'un point de vue

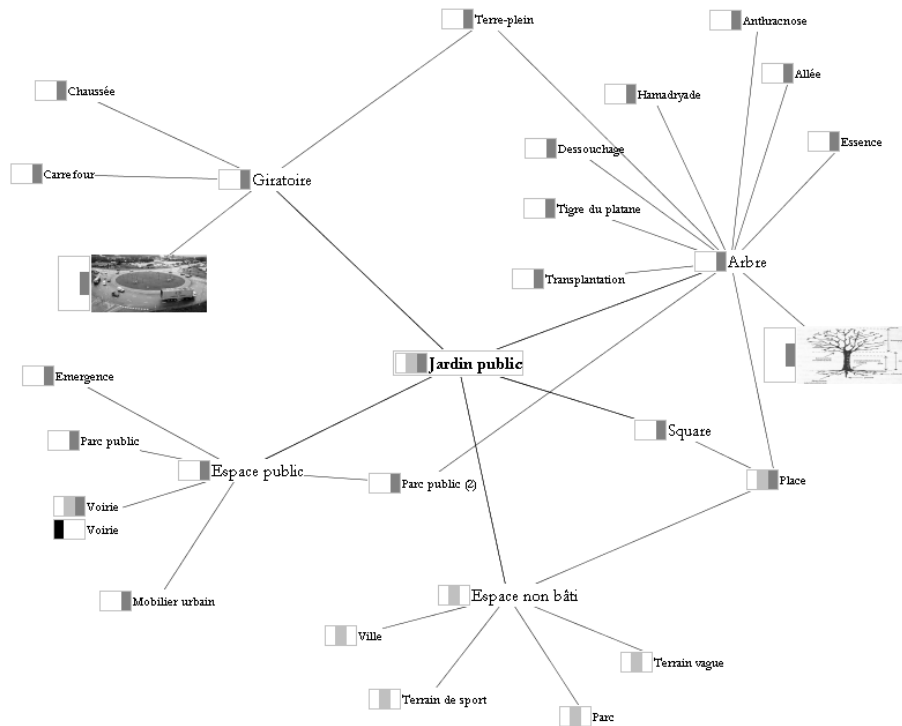
– l'**observation** est l'activité principale de l'expert durant la confrontation. Elle se subdivise en deux modes : l'observation initiale (cf. figure 2), qui vise à naviguer dans les graphes indépendamment les uns des autres, pour en prendre connaissance ; et l'observation expérimentale (cf. figure 3), en cours de confrontation, qui consiste à évaluer l'impact sur ces graphes des interconnexions créées ou supprimées.

– La **désignation** est une activité à destination des points de vue qui ne sont pas des graphes. Elle consiste à sélectionner des motifs du point de vue (que ce soit un

texte, une image ou même un autre support comme une vidéo), motifs qui seront plus tard organisés au sein d'un graphe.

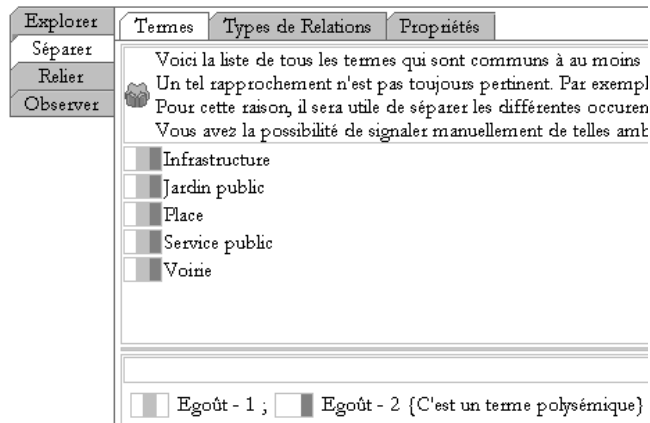
– L'**organisation** interagit avec la désignation : les motifs issus du point de vue sont répertoriés par catégories, selon une structure de graphe orienté acyclique (une structure d'arbre est trop limitée pour exprimer des connaissances d'expert, comme l'explique (Bénel *et al.*, 2002)).

– La **connexion** est un outil habituel d'alignement, et l'interconnexion de points de vue est un moyen puissant de confrontation. Cependant, comme nous l'avons expliqué, le rapprochement de motifs équivalents ou liés par hyp(ér)onymie n'est pas le seul rapprochement qui est fait par les experts. La connexion va donc bien au-delà des considérations de correspondance.



**Figure 3.** Observation locale des points de vue en cours de confrontation

– La **séparation** (voir figure 4) est le second outil nécessaire à l'« alignement étendu » de nos points de vue. Il ne s'agit pas d'une simple opération d'annulation de connexion, mais bien d'un exercice de discernement : il y a des connexions qui semblent évidentes. Il est alors nécessaire de préciser quand ces connexions ne sont pas pertinentes.



**Figure 4.** Séparation de deux termes identiques, ici deux occurrences d'un terme polysémique.

Comme nous l'avons dit, toutes ces activités sont disponibles à tout moment, sans considération de séquence. Ainsi, on peut parfaitement tisser un lien entre deux termes, puis regarder le graphe de voisinage obtenu (voir figure 3), y découvrir un nouveau lien (dans le cas de la figure 3, on pourrait vouloir relier « parc » (gris clair, en bas au milieu) et « parc public » (gris foncé, en bas à gauche), par exemple), tisser ce nouveau lien, regarder à nouveau le graphe etc.

### 5.3. Assistance à la Confrontation

#### 5.3.1. Entités

Nous avons défini un formalisme pivot adapté à l'alignement. Ce formalisme est composé de cinq entités : les nœuds, les liens, les types de nœuds et de liens, et les propriétés des types de liens.

Nos algorithmes d'alignement se concentrent sur trois entités seulement, à savoir les nœuds, les types de liens et les propriétés des types de liens. En effet, les termes liés aux types de nœuds et aux liens ne sont pas utilisés de manière significative dans le cas des points de vue dont nous disposons. Nous ne voulons pas embarrasser l'expert avec des fonctionnalités qui n'ont aucun sens pour lui. D'autre part, il ne faut pas non plus l'embarrasser avec un algorithme d'alignement trop long. Il est cependant difficile d'effectuer un travail précis avec peu de temps. Nous avons résolu ce compromis en proposant trois modes de recherche différents.

### 5.3.2. Recherche exhaustive

Ce premier mode permet d'effectuer un alignement exhaustif sans pour autant bloquer l'interface. Ainsi, l'utilisateur peut profiter du résultat sans avoir dû interrompre son travail. La gestion des priorités entre processus permet de profiter des instants où l'utilisateur n'effectue aucune interaction pour faire le calcul. Cet algorithme utilise l'état des connaissances au moment de l'appel (ce qui a déjà été calculé et ce que l'utilisateur a précisé) pour en proposer une nouvelle itération.

La forme finale de son implémentation n'est pas arrêtée à cette date, pour des questions d'optimisation. Nous présentons donc ici l'idée générale, l'algorithme dans sa version actuelle étant peu efficace s'il est appliqué tel quel (sa complexité est en puissance quatre).

La recherche exhaustive est menée par un algorithme itératif. A chaque itération, les résultats de l'itération précédente sont adaptés suivant ce que l'expert a validé ou refusé. L'utilisation de l'itération précédente nous permet de tirer parti de la structure des graphes sans pour cela générer de référence circulaire. En effet :

- Le calcul de la correspondance entre nœuds tient compte de ce qui a déjà été calculé dans ce domaine, ainsi que des calculs concernant la correspondance entre types de relations. Ainsi, si deux types de relation ont été connectés par l'expert, l'itération suivante pourra rapprocher des nœuds reliés par ce type de relation.
- Le calcul de la correspondance entre type de relations tient compte à la fois des couples de nœuds reliés de part et d'autre (et en particulier de connexions entre certains de ces nœuds), et des propriétés respectives de ces types de relation.
- Le calcul de correspondance entre propriétés tire parti des correspondances entre les types de liens qui ont ces propriétés.

### 5.3.3. Recherche rapide

Cette première fonctionnalité est cependant difficilement envisageable sur de grosses masses de données : un alignement exhaustif pourrait prendre plusieurs jours, temps durant lequel il est peu probable que l'utilisateur laisse l'application en fonctionnement. Nous proposons donc un algorithme plus rapide, nommé dans l'interface recherche rapide, moins précis que le premier, mais apte à effectuer le travail en quelques minutes.

Cet algorithme fait une simple mesure de similarité entre les termes utilisés dans les différents points de vue. Plus précisément, pour chaque couple de points de vue distincts, nous parcourons la liste des termes de part et d'autre correspondant à l'entité choisie (les nœuds, les types de relations entre eux ou les propriétés de ces types de relation). Cet algorithme a donc une complexité quadratique par rapport au nombre de points de vue (il s'agit d'une série arithmétique de raison 1), ainsi que par rapport au nombre de termes.

Le résultat renvoyé est une série de correspondances entre termes de l'entité choisie, chaque correspondance étant assortie d'un coefficient de certitude donné par la mesure de similarité que nous utilisons. Cette mesure de similarité doit donc être normalisée et symétrique. Elle peut varier suivant le corpus de points de vue étudié.

#### 5.3.4. Recherche immédiate

Ce troisième mode de recherche ne diffère du second que par sa couverture : alors que la recherche rapide va parcourir l'ensemble des entités des points de vue, la recherche immédiate utilisera un générateur de nombres aléatoires pour choisir un certain nombre d'entités à examiner, jusqu'à ce que le temps imparti soit écoulé. Ainsi, en choisissant au hasard les parties du graphe à explorer avec l'algorithme précédent, on obtient un certain nombre de mises en relation tout en garantissant à la fois une couverture homogène des données et la tenue d'un délai fixé. Couverture homogène, car les points du graphe explorés sont choisis selon une distribution uniforme. Tenue du délai, parce que les parties choisies sont suffisamment petites pour que leur temps d'exploitation se compte en millisecondes.

#### 5.3.5. Discussion

Chacun des algorithmes a ses avantages et ses inconvénients, qui se déclinent surtout par le compromis entre exhaustivité, qualité et délai d'exécution. Nous préférons laisser le choix à l'utilisateur, maître de son temps, plutôt que de lui imposer un compromis hasardeux qui pourrait ne pas convenir. Comme partout dans notre approche, il s'agit de lui laisser le plus de latitude possible pour qu'il puisse choisir lui-même ses outils.

Nous l'avons déjà évoqué : bien que fort utiles en ce qui concerne le rapprochement de motifs similaires dans le graphe, il va de soi que l'efficacité de ces algorithmes est très faible dans les autres cas (par exemple, détecter des motifs qui ont un lien d'analogie entre eux est très difficile).

### 5.4. Validation par l'expert

Afin de compléter les capacités d'alignement de la machine lors de la confrontation, il est nécessaire, du point de vue de l'expert, de diviser cette activité en deux : ce sont les activités que nous avons présenté sous les termes de connexion et de séparation.

– D'une part, en effet, la machine a une certaine vision, dirigée par ses algorithmes, de correspondances entre motifs des différents points de vue. L'expert seul est à même de valider ces correspondances (ce qui revient à entériner une connexion) ou de les réfuter comme non pertinentes (c'est une séparation).

– D'autre part, il y a des connexions qui ne sont pas détectées par les algorithmes, comme nous l'avons vu. Ces connexions sont faites par l'expert.

– Enfin, il y a des connexions qui sont faites par défaut, par exemple entre des termes identiques. Certaines d’entre elles, non pertinentes, doivent être défaites.

Une correspondance suggérée par la machine est toujours assortie d’un coefficient de certitude compris entre 0 et 100%. Une connexion validée par un expert est considérée comme assortie d’une certitude de 100%. Une séparation effectuée par l’expert, ou une connexion refusée, est assortie d’une certitude de 0.

La combinaison d’opérations élémentaires de connexion et de séparation sur de mêmes motifs reste très complexe à gérer, car nous ne modifions pas les points de vue durant la confrontation. Nous permettons donc l’interconnexion d’éléments séparés, mais pas la séparation d’éléments par ailleurs connectés. Dans ce dernier cas, nous laissons à l’expert le choix entre la destruction des connexions impliquant les termes à séparer (il peut les reconstruire ensuite) et l’abandon de la séparation.

## **6. Conclusion et Perspectives**

En utilisant les possibilités offertes par l’alignement d’ontologies, nous avons donc construit un environnement permettant la confrontation de points de vue, sous réserve que ces points de vue soient exprimés sous forme de graphes. Etant donné la contrainte posée par l’assujettissement de l’environnement à l’expertise de l’utilisateur, nous avons adapté un certain nombre de fonctionnalités aux corpus de points de vue dont nous disposons.

Notre travail actuel se concentre principalement autour de deux axes : d’une part l’implémentation de la recherche exhaustive, pour tirer pleinement partie des potentialités amenées par l’alignement ; d’autre part la modélisation à la volée, basée sur la fonctionnalité de désignation que nous avons exposée. Cela permettra la confrontation de points de vue qui ne sont pas sous forme de graphe. Pour l’heure, cependant, la construction des graphes est faite *via* les plateformes Porphyry et Townology, et nous confrontons donc des graphes déjà construits.

Notre approche présente un certain nombre de limites, qui laissent des questions ouvertes. Par l’utilisation d’un langage pivot, nous nous sommes affranchis de la plupart des problèmes posés par la confrontation dans différents formalismes. Il reste cependant le problème de la confrontation multilingue. Par ailleurs, les formalismes de graphes permettent l’utilisation de ressources multimédia dont l’alignement est difficile. L’aide de l’expert nous permet heureusement de limiter le handicap dans un certain nombre de cas.

## **7. Bibliographie**

Aussenac-Gilles N., « Ontology or meta-model for retrieving scientific reasoning in documents: the Arkeotek project », *Proceedings of the Workshop on Exploring the limits*

*of global models for integration and use of historical and scientific information*, Héraklion, 24-25 octobre 2006.

- Bénel A., « Porphyry au pays des paestans : usages d'un outil d'analyse qualitative de documents par des étudiantes de maîtrise en iconographie grecque », *Actes du colloque international "Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation"*, Albi, 10-14 jul. 2006, Paris, Texto!, vol. 11, n°2, juin 2006, p.182-189.
- Bénel A., Iacovella A., Calabretto S., « Porphyry and Steatite: Software layers for sense makers in humanities », *Proceedings of the workshop on Indexing and Knowledge in Human Sciences*, Nantes, 26-28 Juin 2006, p.72-75.
- Bénel A., Calabretto S., Iacovella A., Pinon J-M., .., « Porphyry 2001: Semantics for scholarly publications retrieval », *Proceedings of ISMIS 2002*, Lyon, 26-29 juin 2002, Heidelberg, Springer Verlag., LNAI, Vol. 2366, p.351-361.
- Breuker J., Van de Velde W., *CommonKads Library for Expertise Modelling*, Amsterdam, IOS Press, 1994.
- El Asri B., Nassar M., Coulette B., « Multiviews Components for Information System Development », *ICEIS 2005*, Miami, 24-28 Mai 2005, USA, INSTICC Press.
- Euzenat J., Shvaiko P., *Ontology Matching*, Heidelberg, Springer Verlag., 2007.
- Gruber, T., « A translation Approach to portable ontology specifications », *Knowledge Acquisition*, London, Academic Press, vol. 5, 1993, pp 199-220.
- Guha R.V., Lenat D.B., « CYC : a midterm report », *AI Magazine*, n° 11, 1990.
- Huzella M., Harzallah M., Trichet F., « OntoMas: a Tutoring System Dedicated to Ontology Matching », *Proceedings of the Third International Conference on Interoperability for Enterprise Software and Applications*, Funchal, 28-30 mars 2007, Heidelberg, Springer Verlag., LNCS, p. 377-388.
- Iacovella A., Bénel A., Pétard X., Helly B., « Corpus scientifiques numérisés : Savoirs de référence et points de vue des experts », *R. T. Pédaugue (Ed.), La redocumentarisation du monde*, Toulouse, Cépaduès, 2006, p. 117-130.
- Levenshtein V.I., « Binary codes capable of correcting deletions, insertions, and reversals », *Soviet Physics Doklady*, 1966, p.707-710.
- Keita A.K., Roussey C., Laurini R., « Un outil d'aide à la construction d'ontologies pré-consensuelles : le projet Towntology », *Actes du XXIVème congrès INFORSID*, Hammamet, 31 Mai-3 Juin 2006, Paris, Hermès, p. 911-926.
- Lahna B., Roudies O., Giraudin J.P., « Une approche multivue pour la conception de systèmes d'information à composants », *Actes du XXIIIème congrès INFORSID*, Grenoble, 24-27 Mai 2005, Paris, Hermès, p. 45-59.
- Piolat A., Vauclair J., « Le processus d'expertise éditoriale avant et après Internet », *Pratiques Psychologiques*, vol. 10, issue 3, 2004, p. 255-272.